Variant Interpretation: A Major Challenge in Applying Genomics to Medicine

Robert L. Nussbaum, MD
Chief Medical Officer, Invitae
Formerly the Holly Smith Professor of Medicine, UCSF

Disclosures

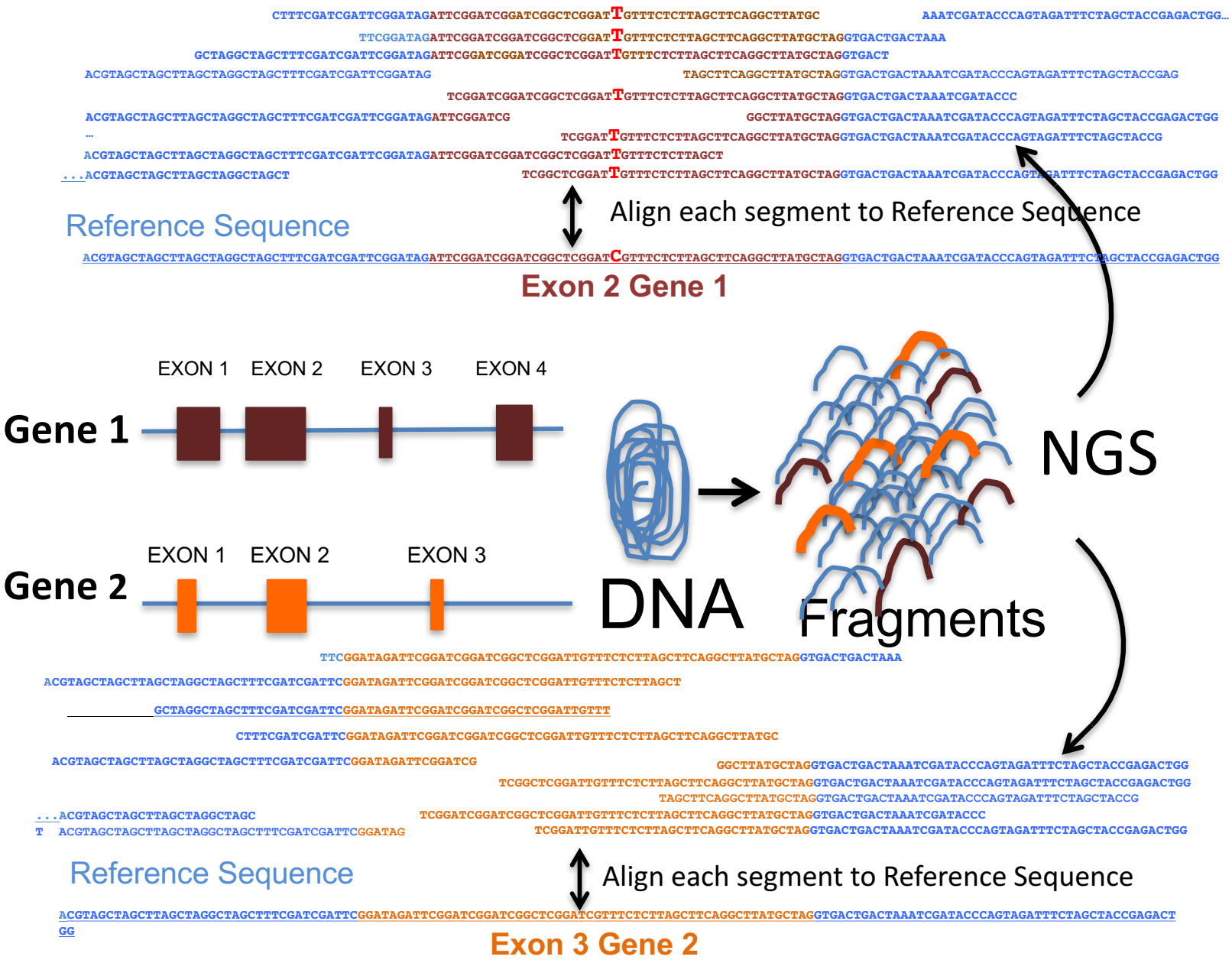
- Officer and Stock Holder at Invitae
- Scientific Advisory Board, Genome Medical
- Chair, Rare Disease Therapeutic Area Scientific Advisory Board, Pfizer

ACCE Framework

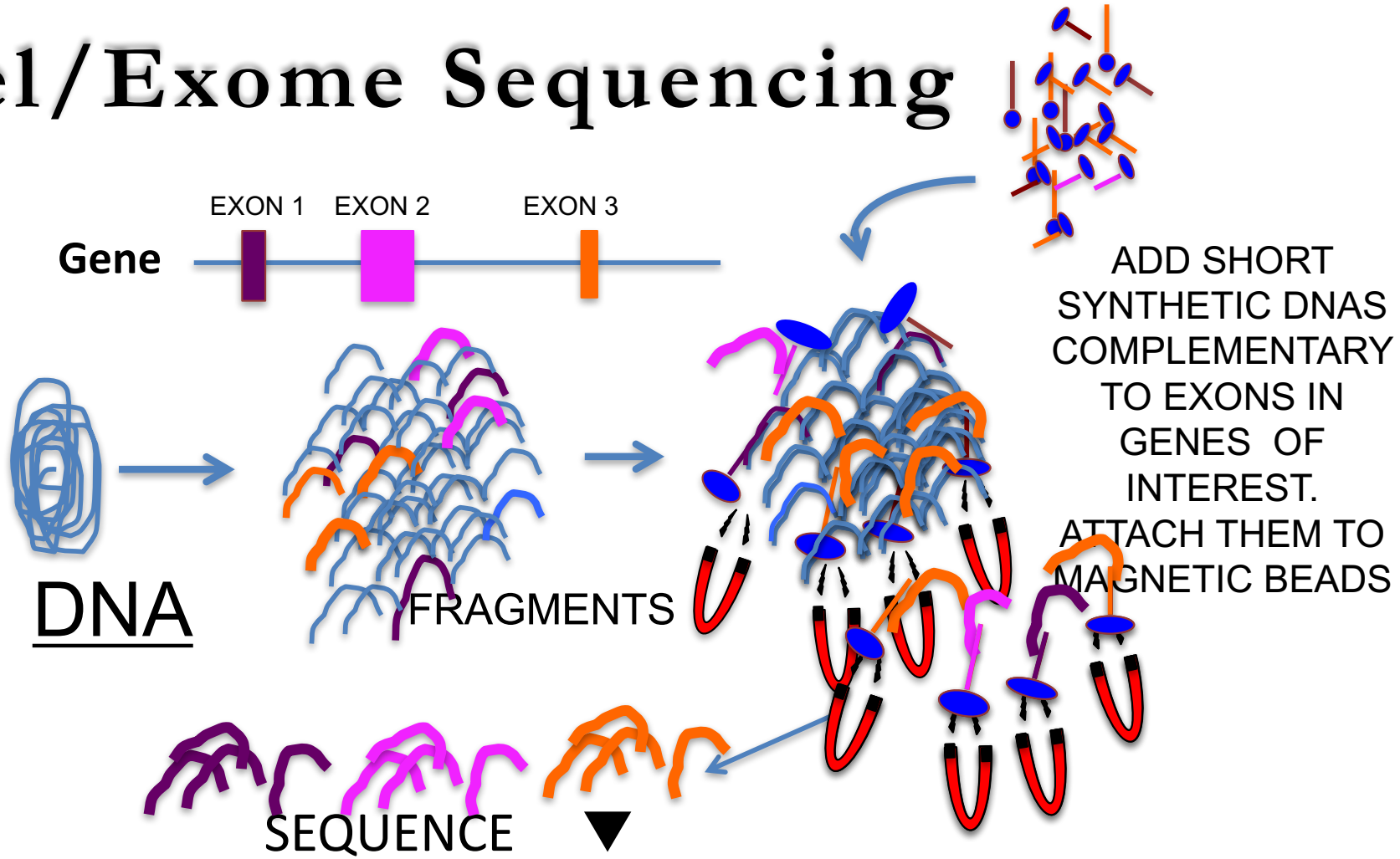
Right Result from the right patient (Sensitivity, Specificity, Accuracy)	***A**nalytic Validity*
Penetrance and Positive and Negative Predictive Values	***C**linical Validity*
Test results are "useful" to patient and doctor Test results "make a difference"	***C**linical Utility*
There is value to society in having test results	***E**thical, **E**conomic Legal, Social Implications*

CDC Office of Public Health Genomics

Whole Genome Sequencing



Panel/Exome Sequencing



TTCGGATAGATTCGGATCGGATCGGCTCGGATTGTTTCTCTTAGCTTCAGGCTTATGCTAGGTGACTGACTAAA
ACGTAGCTAGCTTAGCTAGGCTAGCTTTCGATCGATTCGGATAGATTCGGATCGGATCGGCTCGGATTGTTTCTCTTAGCT
GCTAGGCTAGCTTTCGATCGATTCGGATAGATTCGGATCGGATCGGCTCGGATTGTTT
CTTTCGATCGATTCGGATAGATTCGGATCGGATCGGCTCGGATTGTTTCTCTTAGCTTCAGGCTTATGC AAATCGATACCCAGTAGATTTCTAGCTACCGAGACTGG
ACGTAGCTAGCTTAGCTAGGCTAGCTTTCGATCGATTCGGATAGATTCGGATCG GGCTTATGCTAGGTGACTGACTAAATCGATACCCAGTAGATTTCTAGCTACCGAGACTGG
TCGGCTCGGATTGTTTCTCTTAGCTTCAGGCTTATGCTAGGTGACTGACTAAATCGATACCCAGTAGATTTCTAGCTACCGAGACTGG
ACGTAGCTAGCTTAGCTAGGCTAGCTTTCGATCGATTCGGATAG TAGCTTCAGGCTTATGCTAGGTGACTGACTAAATCGATACCCAGTAGATTTCTAGCTACCGAG
TCGGATCGGATCGGCTCGGATTGTTTCTCTTAGCTTCAGGCTTATGCTAGGTGACTGACTAAATCGATACCC
...ACGTAGCTAGCTTAGCTAGGCTAGCT
TCGGATTGTTTCTCTTAGCTTCAGGCTTATGCTAGGTGACTGACTAAATCGATACCCAGTAGATTTCTAGCTACCGAGACTGG

Align each segment to Reference Sequence

ACGTAGCTAGCTTAGCTAGGCTAGCTTTCGATCGATTCGGATAGATTCGGATCGGATCGGCTCGGATCGTTTCTCTTAGCTTCAGGCTTATGCTAGGTGACTGACTAAATCGATACCCAGTAGATTTCTAGCTACCGAGACTGG

Reference Sequence

Exon 3

Clinical Validity

Positive Predictive Value: In people with + test ➔Disease present or increased risk?

Negative Predictive Value: In people with - test ➔Disease absent or population risk?

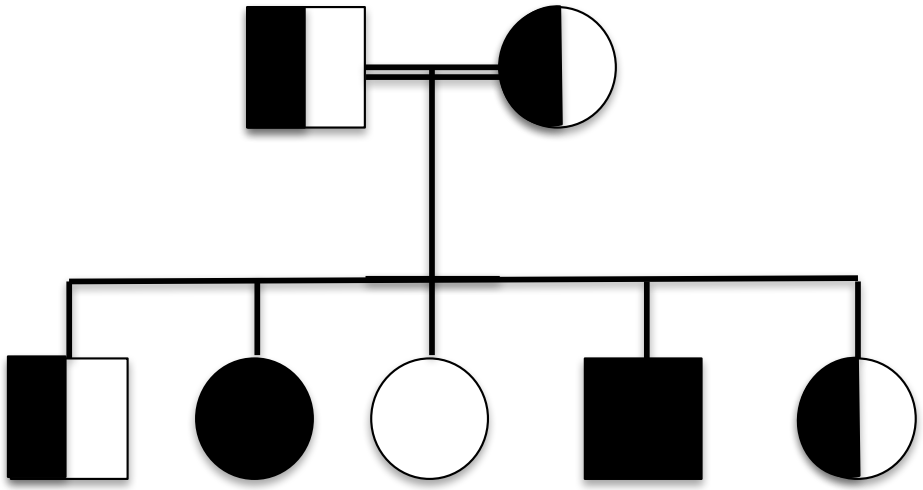
When You Compare an Individual's Whole Genome Sequence to the Reference, What Do You Find?

	Single Nucleotide Variants	Insertion/Deletion
Total Number	3,500, 000	500,000
Number within Genes	1,340,000	120,000
Number in Exons	47,000	5,800
Number in Coding Exons	20,000	470
New Stop Codon (Nonsense Mutation)	82	-
Frame Shift	-	255
Changes an Amino Acid	10,500	12
No Amino Acid Change	9,300	-

Variant Interpretation

- Gene-Disease Relationship
- Variant-Disease Relationship

Two Siblings with Infantile Epilepsy



Synaptojanin I
c.773 G>A
p.Arg258Gln

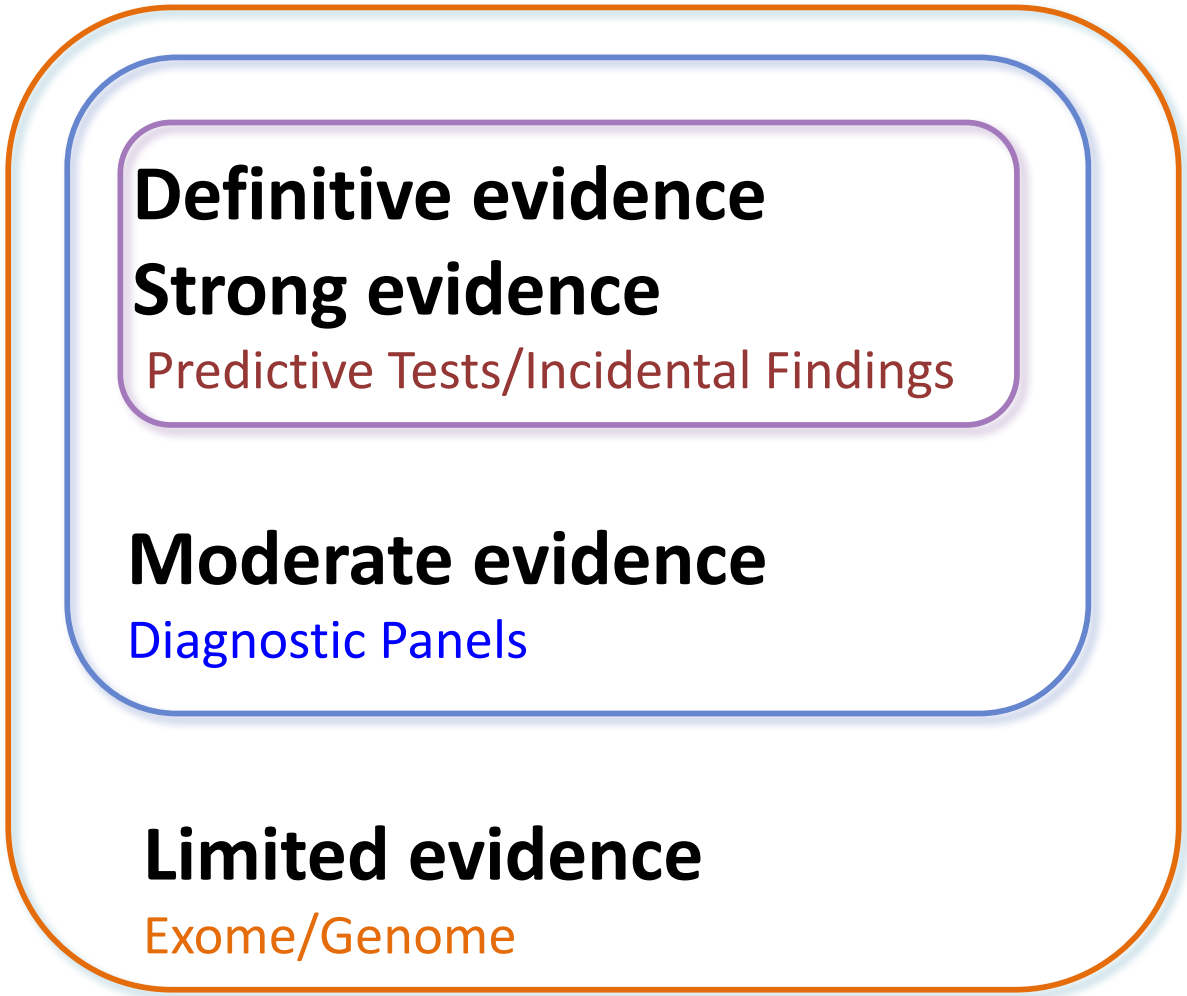
Mutation in the NH2 -terminal Sac1-like inositol phosphatase domain of polyphosphoinositide phosphatase synaptojanin 1 (SYNJ1)
Gene product is implicated in the regulation of endocytic traffic at synapses

Gene-Disease Evidence Levels

Evidence Level	Evidence Description
DEFINITIVE	The role of this gene in this particular disease has been repeatedly demonstrated in both the research and clinical diagnostic settings, and has been upheld over time (in general, at least 3 years). No valid evidence has emerged that contradicts the role of the gene in the specified disease.
STRONG	There is **strong** evidence by at least two independent studies to support a causal role for this gene in this disease, such as: •Strong statistical evidence demonstrating an excess of pathogenic variants[1] in affected individuals as compared to appropriately matched controls •Multiple pathogenic variants within the gene in unrelated probands with several different types of supporting experimental data[1]. The number and type of evidence might vary (eg. fewer variants with stronger supporting data, or more variants with less supporting data) In addition, no valid evidence has emerged that contradicts the role of the gene in the noted disease.
MODERATE	There is **moderate** evidence to support a causal role for this gene in this disease, such as: •At least 3 unrelated probands with pathogenic variants[1] within the gene with some supporting experimental data. The role of this gene in this particular disease may not have been independently reported, but no valid evidence has emerged that contradicts the role of the gene in the noted disease.
LIMITED	There is **limited** evidence to support a causal role for this gene in this disease, such as: •Fewer than three observations of a pathogenic variant[1] within the gene •Multiple variants reported in unrelated probands but *without* sufficient evidence for pathogenicity per 2014 ACMG criteria
NO EVIDENCE	No evidence reported for a causal role in disease.
DISPUTED	Valid evidence of approximate equivalent weight exists both supporting and refuting a role for this gene in this disease.
EVIDENCE AGAINST	Evidence refuting the role of the gene in the specified disease has been reported and significantly outweighs any evidence supporting the role.

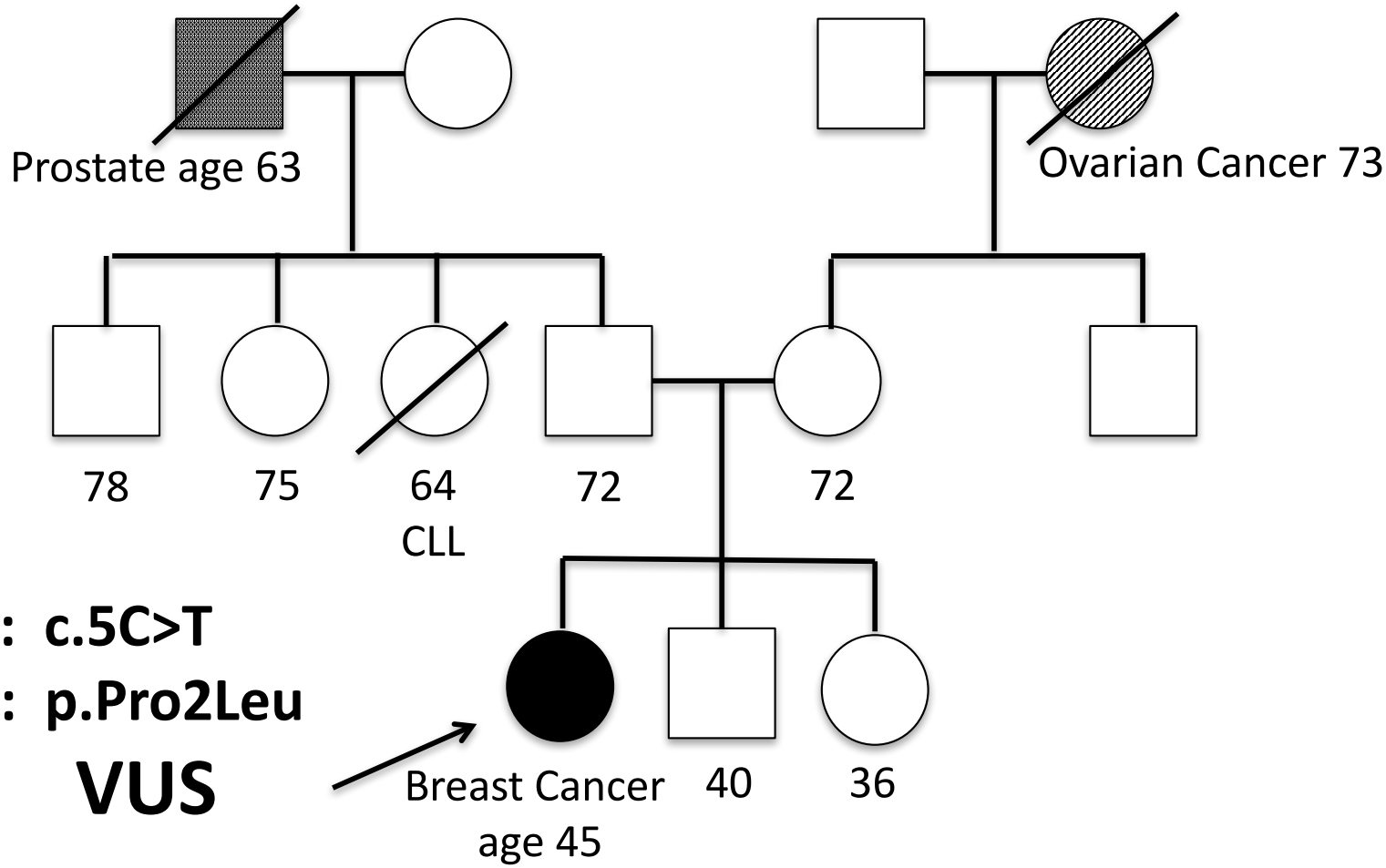
http://clinicalgenome.org/data-curation/clinical-validity/

What Evidence is Required to Include a Gene In a Clinical Report? Depends on the Purpose of the Report



ClinGen

45 Year Old Woman with Breast Cancer



***BRCA2*: c.5C>T**

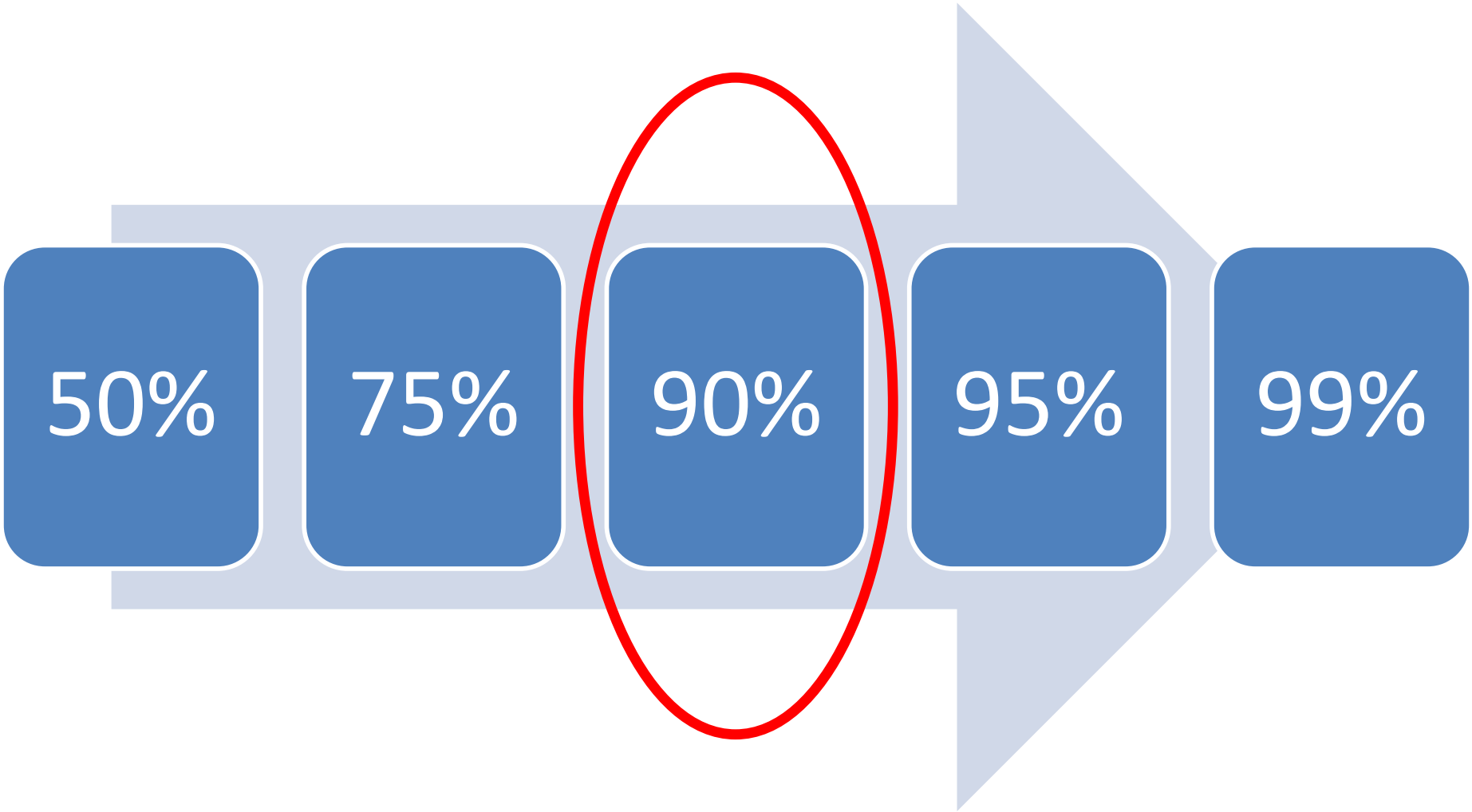
BRCA2: p.Pro2Leu

VUS

ACMG Variant Interpretation Categories

- Pathogenic
- Likely Pathogenic
- Variant of Uncertain Significance
- Likely Benign
- Benign

What is the meaning of ‘Likely’?



	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very Strong**
Population Data	MAF frequency is too high for disorder **OR** observation in controls inconsistent with disease penetrance			Absent or appropriately rare in public databases	Prevalence in affecteds statistically increased over controls	
Computational Data		Multiple lines of computational evidence suggest no impact on gene /gene product Type of variant does not fit known mechanism of disease	Multiple lines of computational evidence support a deleterious effect on the gene /gene product	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before In-frame indels in a non-repeat region Stop-loss variants	Same amino acid change as an established pathogenic variant	**Truncating variant in a gene where LOF is a known mechanism of disease**
Functional Data	Well-established functional studies show no deleterious effect	In-frame indels in a repetitive region without a known function[7]	Missense in gene with low rate of benign missense variation and pathogenic missenses common	Located in a mutational hot spot and/or known functional domain	Well-established functional studies show a deleterious effect	
Segregation Data	Non-segregation with disease		Co-segregation with disease in multiple affected family members	Co-segregation with disease in multiple affecteds in multiple families		
De novo Data				*De novo* (without paternity & maternity confirmed)	*De novo* (paternity & maternity confirmed)	
Allelic Data		Dominants: Observed in *trans* with a pathogenic variant Observed in *cis* with a pathogenic variant		For recessive disorders, detected in *trans* with a pathogenic variant		
Other Database		Reputable database = benign	Reputable database = pathogenic			
Other Data		Found in case with an alternate cause	Patient's phenotype or FH matches gene			

The Scoring Rules for Classification

Pathogenic

1 Very Strong *AND*

1 Strong *OR*

≥2 (Moderate *OR* Supporting)

2 Strong

1 Strong *AND*

≥3 Moderate *OR*

≥2 Moderate and 2 Supporting *OR*

≥1 Moderate and 4 Supporting

Likely Pathogenic

1 Very strong or Strong *AND*

≥1 Moderate *OR*

≥2 Supporting

≥3 Moderate

≥2 Moderate *AND* 2 Supporting

≥1 Moderate *AND* 4 Supporting

Benign

1 Stand Alone *OR*

≥ 2 Strong

Likely Benign

1 Strong and ≥1 Supporting *OR*

>2 Supporting

Uncertain Significance

If other criteria are unmet or arguments for benign and pathogenic are equal in strength

CSER Interpretation Bake-Off v2.0: 99 Variants x 9 Labs

- 99 variants were considered, representing all categories (pathogenic, likely pathogenic, uncertain significance (VUS), likely benign, and benign).
- 9 were classified by all 9 labs, 90 variants were classified by 3-4 labs (mean of 3.01) using both the lab's own classification system and also the ACMG guidelines.
- We evaluated both intra-laboratory and inter-laboratory differences among variant classifications using the labs' criteria vs. adopting ACMG criteria.

Data from Gail Jarvik

Bake Off V 2.0

		Lab Class					
		P	LP	VUS	LB	B	Total
ACMG Class	P	62	8	5	0	0	75
	LP	12	55	4	0	0	71
	VUS	2	6	94	17	4	123
	LB	0	0	3	34	7	44
	B	0	0	0	4	30	34
Total		76	69	106	55	41	**347**

79% Identical

ClinVar Discordance – HOT TOPIC

The good, the bad and the ugly

BRCA1/2 data concordance data in ClinVar (May 2016

1. Analysis was limited to data that met objective criteria:
   - Submitted by established clinical labs,
   - Labs had >200 BRCA1/2 classifications in ClinVar,
   - Entries <5 years old

2. Comparisons considered only differences that would significantly change management decisions under current guidelines
   (Pathogenic/Likely Pathogenic versus VUS/Likely Benign/Benign)

Pairwise Concordance by Submitter to ClinVar (Clinically Actionable versus Not Clinically Actionable)

	Ambry	Invitae	GeneDx	Counsyl	CHEO	Emory
SCRP/Myriad	98.7 1018/1031 (97.9–99.3)	99.0 824/832 (98.2–99.5)	99.3 610/614 (98.5–99.8)	99.4 177/178 (97.4–100)	98.0 145/148 (94.7–99.4)	97.2 106/109 (92.8–99.2)
Ambry		99.3 1052/1059 (98.7–99.7)	99.6 777/780 (99.0–99.9)	99.6 223/224 97.9–100)	98.3 176/179 (95.6–99.5)	98.8 161/163 (96.1–99.7)
Invitae			99.7 664/666 (99.0–99.9)	98.7 220/223 (96.5–99.6)	98.3 177/180 (95.6–99.5)	98.7 151/153 (95.9–99.7)
GeneDx				99.5 220/221 (97.9–100)	97.9 138/141 (94.4–99.4)	99.3 149/150 (96.9–100)
Counsyl					100.0 82/82 (97.0–100)	100.0 105/105 (97.6–100)
CHEO						98.3 57/58 (92.2–99.9)

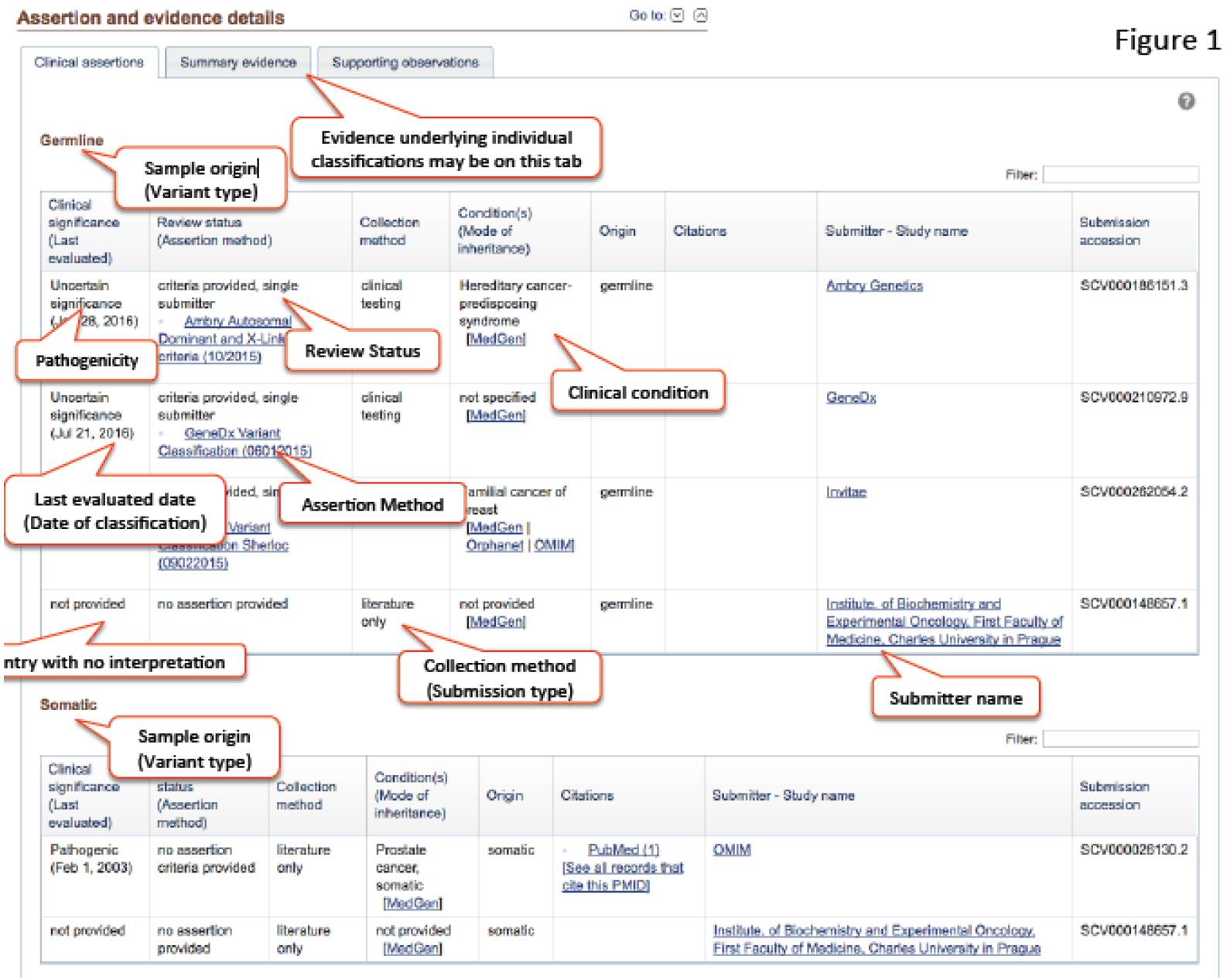
Abbreviations: CHEO, Children's Hospital of Eastern Ontario; SCRP, Sharing Clinical Reports Project.

What is Responsible for Discordance?

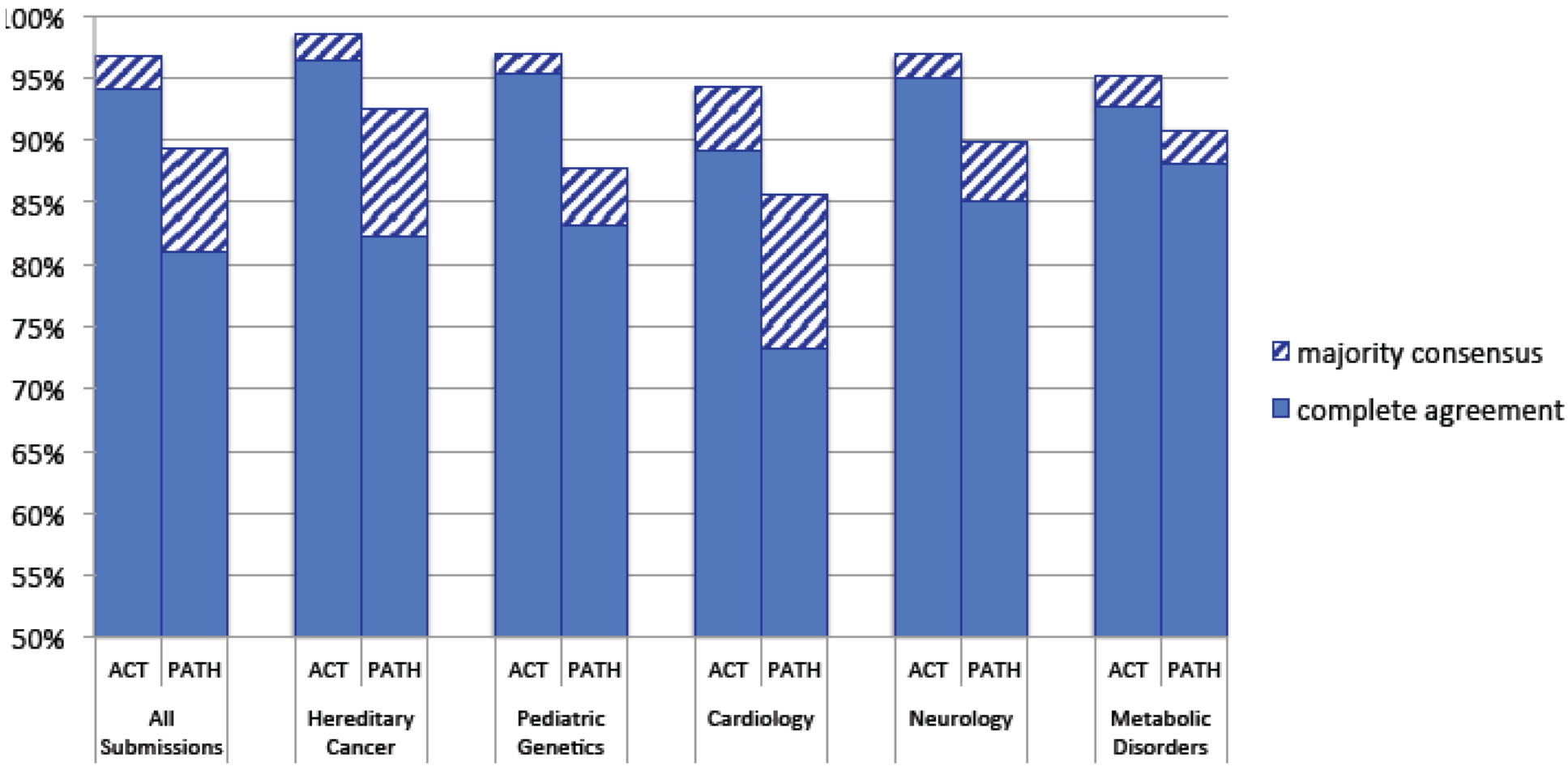
- We evaluated ALL pathogenicity assessments in the ClinVar Sept 2016.
- We included all unique variants from genes Invitae currently offers with at least 2 classifications submitted by established clinical laboratories including data from Myriad Genetics submitted via the Sharing Clinical Reports Project (SCRP).
- 38,011 total classifications of 14,802 unique variants (averaging 2.56 classification per variant) from 520 genes.

ClinVar Entries

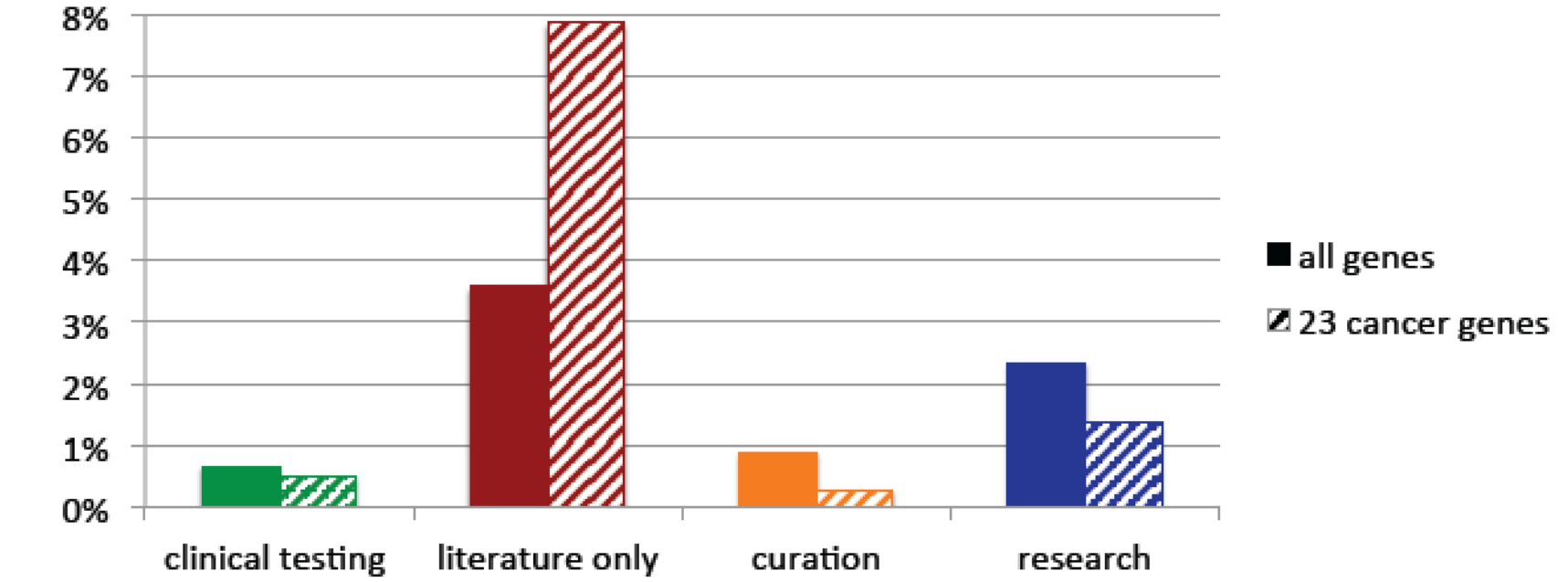
Figure 1

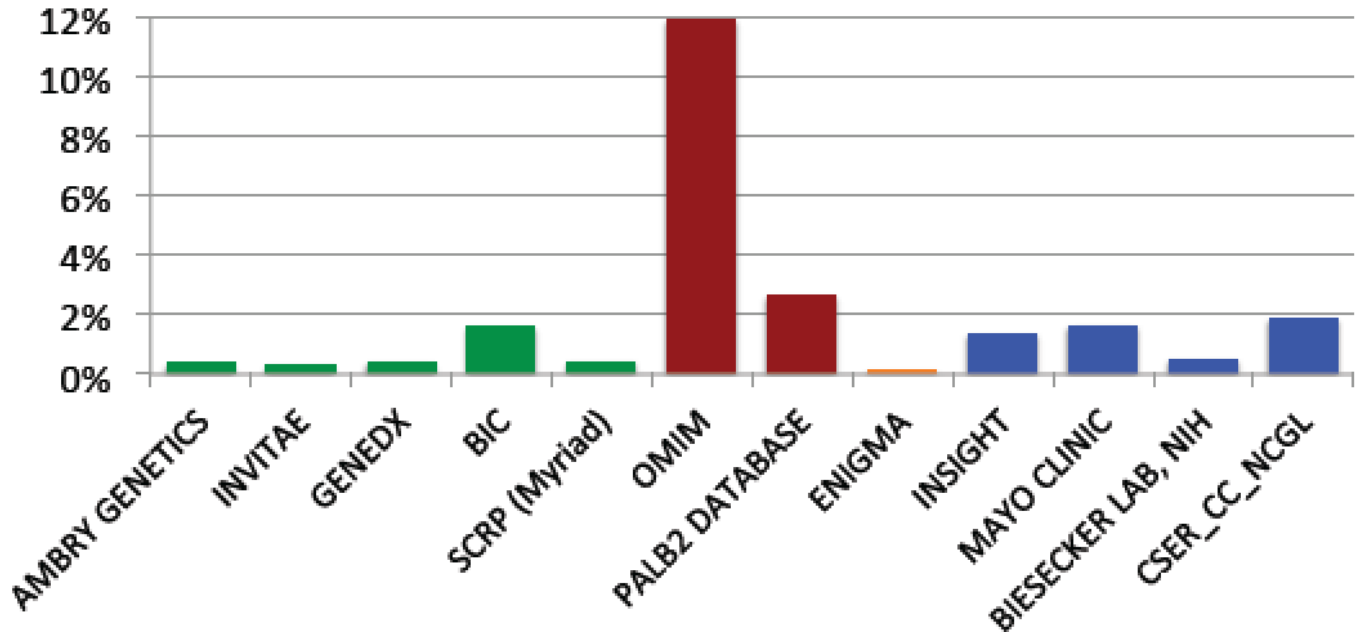


Concordance in Actionability and Pathogenicity Interpretations

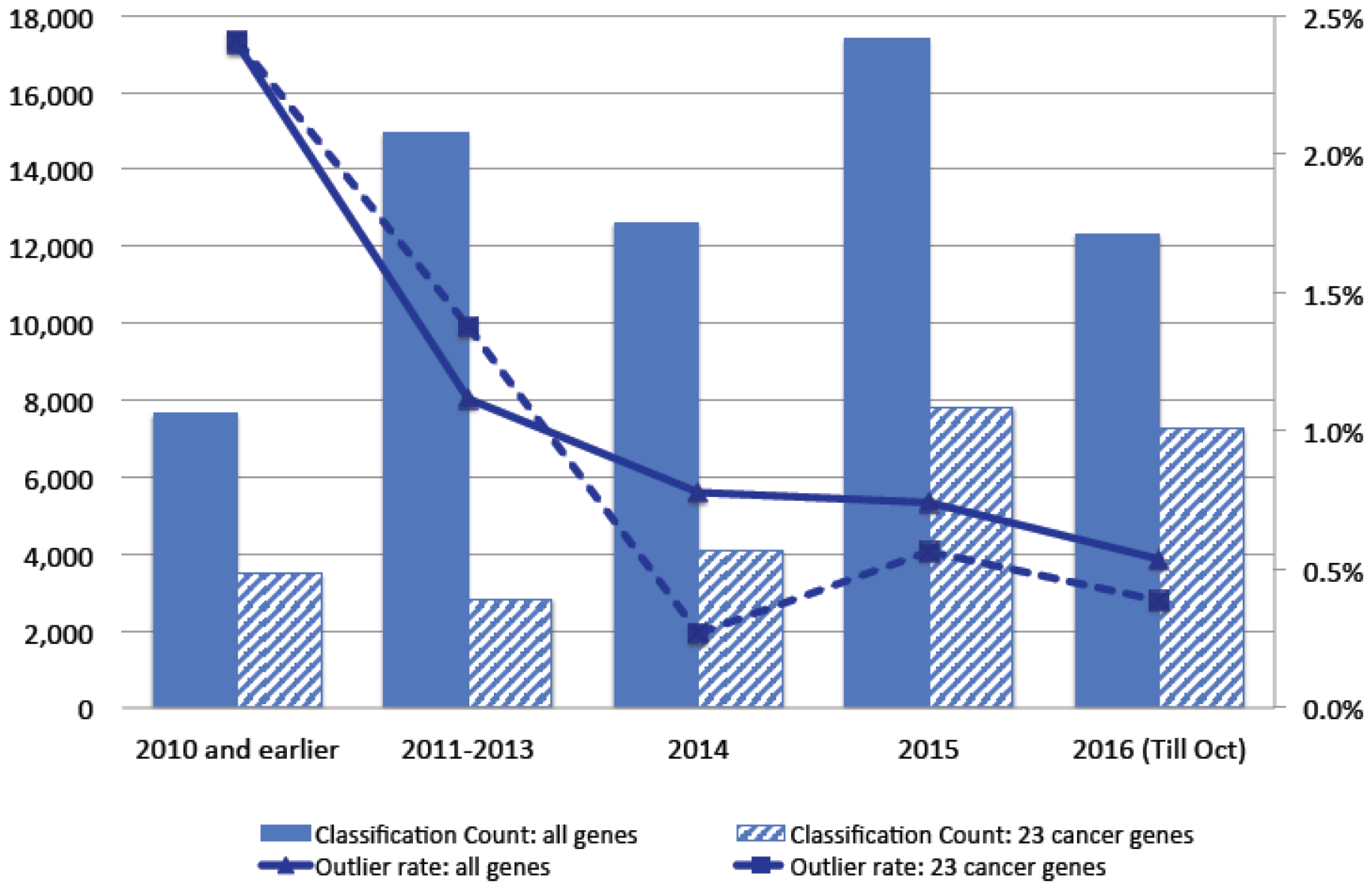


Concordance by Source or ClinVar Submission

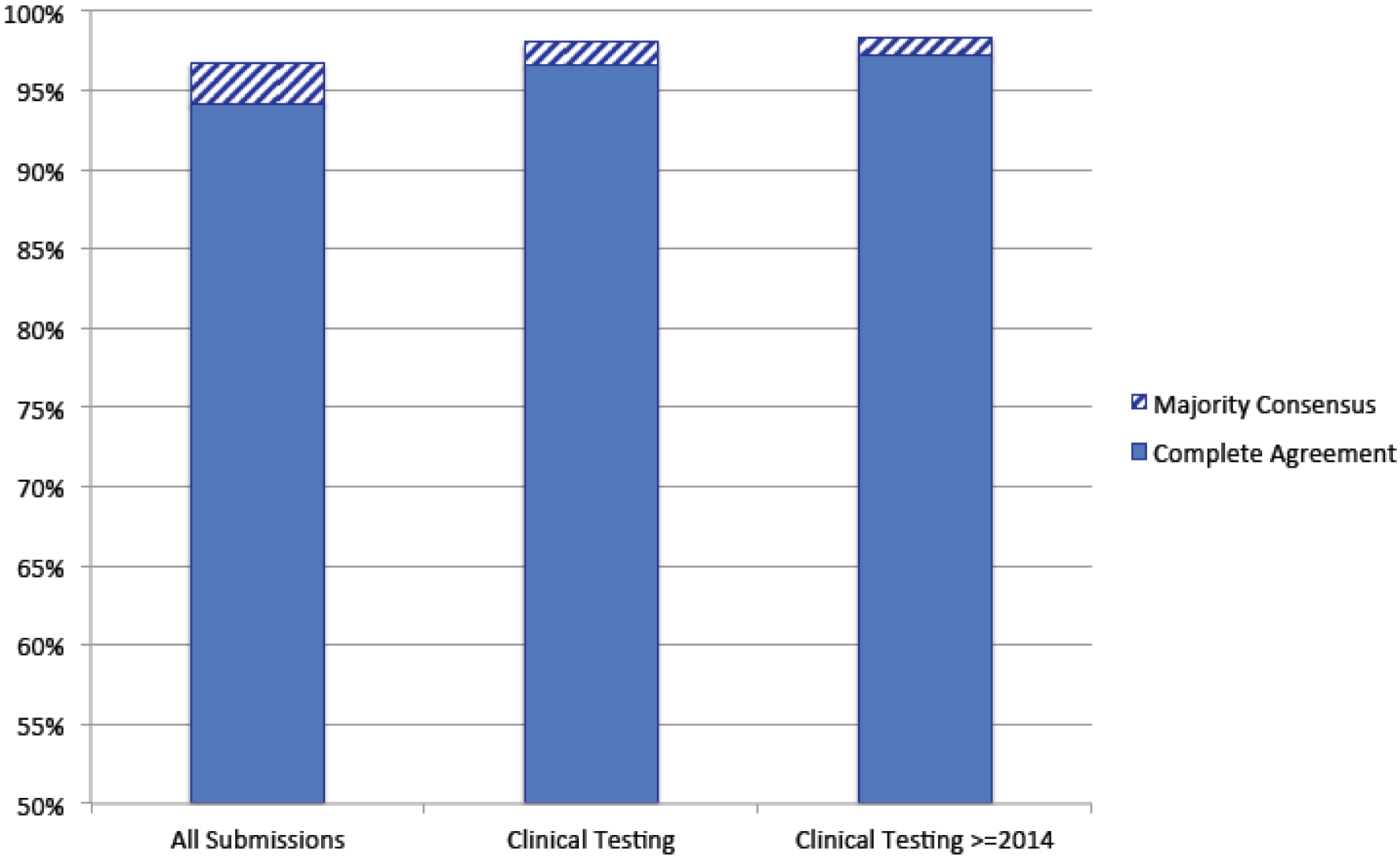




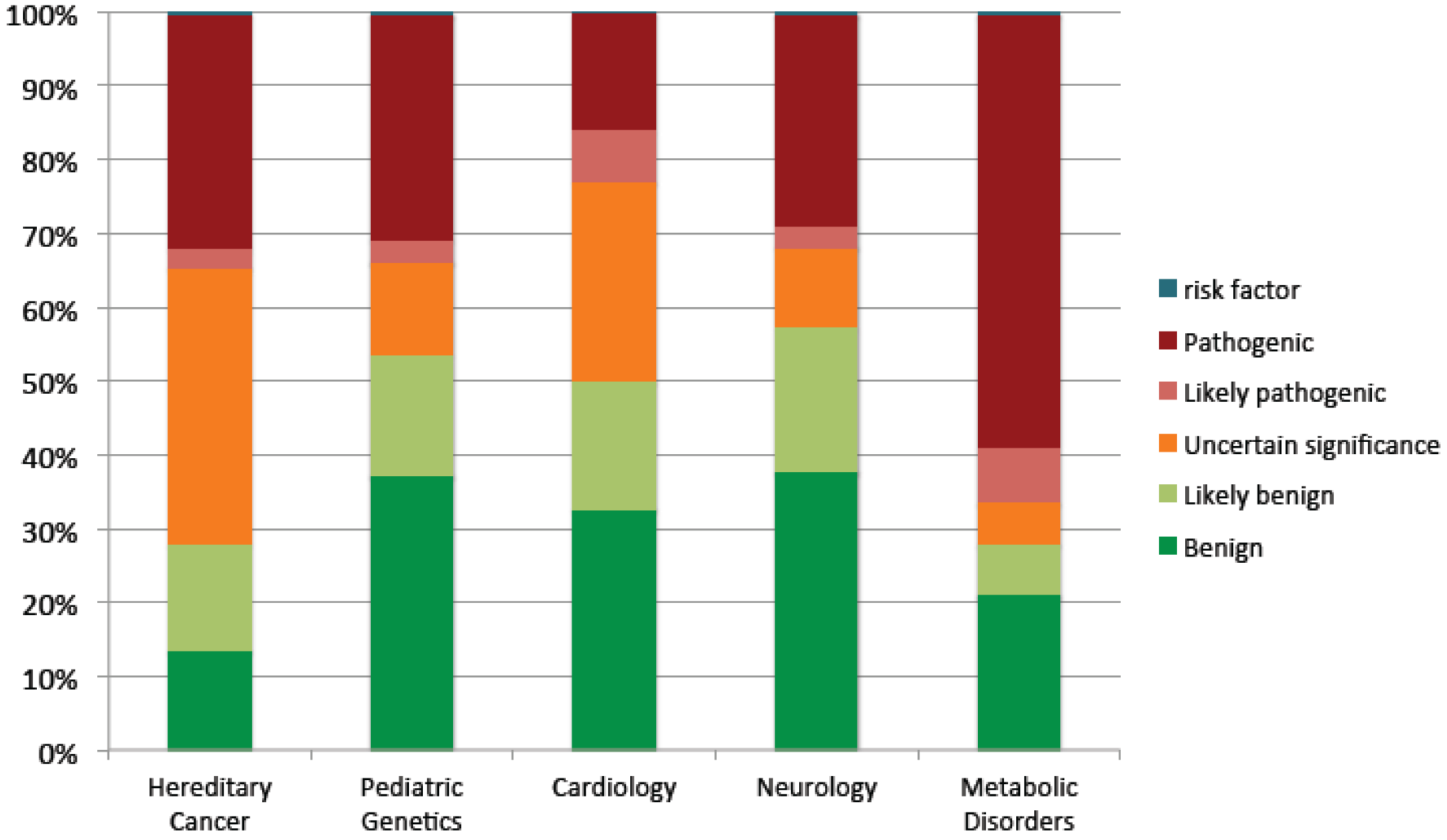
Concordance by Date of Submission



Concordance by Category and Date



Variant Interpretation by Clinical Area



	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very Strong**
Population Data	MAF frequency is too high for disorder **OR** observation in controls inconsistent with disease penetrance			Absent or appropriately rare in public databases	Prevalence in affecteds statistically increased over controls	
Computational Data		Multiple lines of computational evidence suggest no impact on gene /gene product Type of variant does not fit known mechanism of disease	Multiple lines of computational evidence support a deleterious effect on the gene /gene product	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before In-frame indels in a non-repeat region Stop-loss variants	Same amino acid change as an established pathogenic variant	**Truncating variant in a gene where LOF is a known mechanism of disease**
Functional Data	Well-established functional studies show no deleterious effect	In-frame indels in a repetitive region without a known function[7]	Missense in gene with low rate of benign missense variation and pathogenic missenses common	Located in a mutational hot spot and/or known functional domain	Well-established functional studies show a deleterious effect	
Segregation Data	Non-segregation with disease		Co-segregation with disease in multiple affected family members	Co-segregation with disease in multiple affecteds in multiple families		
De novo Data				*De novo* (without paternity & maternity confirmed)	*De novo* (paternity & maternity confirmed)	
Allelic Data		Dominants: Observed in *trans* with a pathogenic variant Observed in *cis* with a pathogenic variant		For recessive disorders, detected in *trans* with a pathogenic variant		
Other Database		Reputable database = benign	Reputable database = pathogenic			
Other Data		Found in case with an alternate cause	Patient's phenotype or FH matches gene			

ACCE Framework

Right Result from the right patient (Sensitivity, Specificity, Accuracy)	***Analytic Validity***
Penetrance and Positive and Negative Predictive Values	***Clinical Validity***
Test results are "useful" to patient and doctor Test results "make a difference"	***Clinical Utility***
There is value to society in having test results	***Ethical, Economic Legal, Social Implications***

CDC Office of Public Health Genomics

What do we Mean by Clinical Utility?